Introduction

This work is devoted to the study of the *scaled gradient projection (SGP) method with non-monotone line search* to solve general constrained convex optimization problems as follows

$$\min\{f(x) : \ x \in C\}, \tag{1}$$?{eq:OptP}?

where C is a closed and convex subset of $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. Denotes by $f^* := \inf_{x \in C} f(x)$ the optimal value of (1) and by Ω^* its solution set, which we will assume to be nonempty unless the contrary is explicitly stated. Problem (1) is a basic issue of constrained optimization, which appears very often in various areas, including finance, machine learning, control theory, and signal processing, see for example [21, 22, 36, 47, 51, 63]. Recent problems considered in most of these areas, the datasets are large or high-dimensional and their solutions need to be approximated quickly with a reasonably accuracy. It is well known that SGP method with nonmonotone line search is among those that are suitable for this task, as will be explained below.

The gradient projection method is what first comes to mind when we start from the ideas of the classic optimization methods in an attempt to deal with problem (1). In fact, this method is one of the oldest known optimization methods to solve (1), the study of its convergence properties goes back to the works of Goldstein [40] and Levitin and Polyak [50]. After these works, several variants of it have appeared over the years, resulting in a vast literature on the subject, including [10, 11, 12, 34, 36, 43, 48, 57, 69]. Additional reference on this subject can be found in the recent review [18] and references therein. Among all the variants of the gradient projection method, the scaled version has been especially considered due to the flexibility provided in efficient implementations of the method; see [14, 5, 17, 19, 20]. In addition, its simplicity and easy implementation has attracted the attention of the scientific community that works on optimization over the years. This method usually uses only first-order derivatives, which makes it very stable from a numerical point of view and therefore quite suitable for solving large-scale optimization problems, see [53, 54, 63, 64]. At each current iteration, SGP method moves along the direction of the negative scaled gradient, and then projects the obtained point onto the constraint set. The current iteration and such projection define a feasible descent direction and a line search in this direction is performed to define the next iteration. In this way, the performance of the method is strongly related to each of the steps we have just mentioned. In fact, the scale matrix and the step size towards the negative scaled gradient are freely selected in order to improve the performance of SGP method but without increasing the cost of each iteration. Strategies for choosing both have their origins in the study of gradient method for unconstrained optimization, papers addressing this issues include but not limited to [7, 19, 27, 28, 30, 37, 71, 26, 50]. It is worth mentioning that, for suitable choices of the scale matrix and the step size, SGP merges into the well known *spectral gradient method* extensively studied in [14, 13]. More details about selecting step sizes and scale matrices can be found in the recent review [18] and references therein.

In this paper, we are particularly interested in the main stages that make up the SGP method, namely, in the projection calculation and in the line search employed. It is well known that the mostly computational burden of each iteration of the SGP method is in the calculation of the projection. Indeed, the projection calculation requires, at each iteration, the solution of a quadratic problem restricted to the feasible set, which can lead to a substantial increase in the cost per iteration if the number of unknowns is large. For this reason, it may not be justified to carry out exact projections when the iterates are far from the solution of the problem. In order to reduce the computational effort spent on projections, inexact procedures that become more and more accurate when approaching the solution, have been proposed, resulting in more efficient methods; see for exemple [14, 17, 39, 42, 62, 66, 59]. On the other hand, nonmonotone searches can improve the probability of finding an optimal global solution, in addition to potentially improving the speed of convergence of the method as a whole, see for example [25, 56, 65]. The concept of nonmonotone line search, that we will use here as a synonym for inexact line search, have been proposed first in [46], and later a new nonmonotone search was proposed in [70]. After these papers others nonmonotone searches appeared, see for example [3, 52]. In [61], an interesting general framework for nonmonotone line search was proposed, and more recently modifications of it have been presented in [44, 45].

The purpose of the present paper is to present an inexact version of the SGP method, which is inexact in two sense. First, using a version of scheme introduced in [14] and also a variation of the one appeared in [66, Example 1], the inexact projection onto the feasible set is computed allowing an appropriate relative error tolerance. Second, using the inexact conceptual scheme for the line search introduced in [45, 61], a step size is computed to define the next iteration. More specifically, initially we show that the feasible inexact projection of [14] provides greater latitude than the projection of [66, Example 1]. In the first convergence result presented, we show that the SGP method using the projection proposed in [14] preserves the same partial convergence result as the classic method, that is, we prove that every accumulation point of the sequence generated by the SGP method is stationary for problem (1). Then, considering the inexact projection of [66, Example 1], and under mild assumptions, we establish full asymptotic convergence results and some complexity bounds. The presented analysis of the method is done using the general nonmonotone line search scheme introduced in [45]. In this way, the proposed method can be adapted to several line searches and, in particular, will allow obtaining several known versions of the SGP method as particular instances, including [10, 14, 48, 68]. Except for the particular case when we assume that the SGP method employs the nonmonotone line search introduced by [46], all other asymptotic convergence and complexity results are obtained without any assumption of compactness of the sub-level sets of the objective function. Finally, it is worth mentioning that the complexity results obtained for the SGP method with a general nonmonotone line search are the same as in the classic case when the usual Armijo search is employed, namely, the complexity bound $\mathcal{O}(1/\sqrt{k})$ is unveil for finding ϵ-stationary points for problem (1) and, under convexity on f, the rate to find a ϵ-optimal functional value is $\mathcal{O}(1/k)$.

In Section 1, some notations and basic results used throughout the paper is presented. In particular, Section 2 is devoted to recall the concept of relative feasible inexact projection and some new properties about this concept are presented. Section **??** describes the SGP method with a general nonmonotone line search and some particular instances of it are presented. Partial asymptotic convergence results are presented in Section **??**. Section **??** presents a full convergence result and iteration-complexity bounds. Some numerical experiments are provided in Section **??**. Finally, some concluding remarks are made in Section **??**.

Chapter 1

Preliminaries and basic results

In this chapter, we introduce some notation and results used throughout our presentation. First we consider the index set $\mathbb{N} := \{0, 1, 2, \ldots\}$, the usual inner product $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^n$, and the associated Euclidean norm $\| \cdot \|$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and $C \subseteq \mathbb{R}^n$. The gradient ∇f of f is said to be *Lipschitz continuous* in C with constant $L > 0$ if $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$, for all $x, y \in C$. Combining this definition with the fundamental theorem of calculus, we obtain the following result whose proof can be found in [12, Proposition A.24].

Lemma 1.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and $C \subseteq \mathbb{R}^n$. Assume that ∇f is Lipschitz continuous in C with constant $L > 0$. Then,*

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq (L/2)\|x - y\|^2,$$

for all $x, y \in C$.

Assume that C is a convex set. The function f is said to be *convex* on C, if

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle,$$

for all $x, y \in C$. We recall that a point $\bar{x} \in C$ is a *stationary point* for problem (1) if

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0, \qquad \forall\, x \in C. \tag{1.1}$$?{eq:StatPoi

Consequently, if f is a convex function on C, then (1.1) implies that $\bar{x} \in \Omega^*$. We will now present some useful concepts for the analysis of the sequence generated by the scaled gradient method, for more details, see [24]. For that, let D be a $n \times n$ positive definite matrix and $\| \cdot \|_D : \mathbb{R}^n \to \mathbb{R}$ be the norm defined by

$$\|d\|_D := \sqrt{\langle Dd, d \rangle}, \quad \forall d \in \mathbb{R}^n. \tag{1.2}$$?{def:normaD

For a fixed constant $\mu \geq 1$, *denote by $\mathcal{D}_\mu$ the set of symmetric positive definite matrices $n \times n$ with all eigenvalues contained in the interval* $[\frac{1}{\mu}, \mu]$. The set $\mathcal{D}_\mu$ is compact. Moreover, for each $D \in \mathcal{D}_\mu$, it follows that D^{-1} also belongs to $\mathcal{D}_\mu$. Furthermore, due to $D \in \mathcal{D}_\mu$, by (1.2), we obtain

$$\frac{1}{\mu}\|d\|^2 \leq \|d\|_D^2 \leq \mu\|d\|^2, \qquad \forall d \in \mathbb{R}^n. \tag{1.3}$$?{eq:pnv}?

Let us recall the the concept of sequence quasi-Fejér monotone to a set, introduced in [24].

Definition 1.2. *Let $(y^k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$ and $(D_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{D}_\mu$. The sequence $(y^k)_{k\in\mathbb{N}}$ is said to be quasi-Fejér monotone to a set $W \subset \mathbb{R}^n$ with respect to $(D_k)_{k\in\mathbb{N}}$ if, there exists a sequence $(\eta_k)_{k\in\mathbb{N}} \subset [0,+\infty)$ such that $\sum_{k\in\mathbb{N}} \eta_k < \infty$ and for all $w \in W$, there exists a sequence $(\epsilon_k)_{k\in\mathbb{N}} \subset [0,+\infty)$ such that $\sum_{k\in\mathbb{N}} \epsilon_k < \infty$, and*

$$\|y_{k+1} - w\|^2_{D_{k+1}} \leq (1+\eta_k)\|y^k - w\|^2_{D_k} + \epsilon_k,$$

for all $k \in \mathbb{N}$.

The following lemma is useful to study the quasi-Fejér monotone sequence, its prove can be found in [58, Lemma 2.2.2].

Lemma 1.3. *Let $(\alpha_k)_{k\in\mathbb{N}}$, $(\eta_k)_{k\in\mathbb{N}}$ and $(\epsilon_k)_{k\in\mathbb{N}}$ be a sequences in $[0,+\infty)$ such that $\sum_{k\in\mathbb{N}} \eta_k < \infty$ and $\sum_{k\in\mathbb{N}} \epsilon_k < \infty$. Assume that $\alpha_{k+1} \leq (1+\eta_k)\alpha_k + \epsilon_k$, for all $k \in \mathbb{N}$. Then, $(\alpha_k)_{k\in\mathbb{N}}$ converges.*

The main property of quasi-Fejér monotone sequences is stated in the following. Its proof can be found in [24, Proposition 3.2 and Theorem 3.3]. For sake of completeness, we include it here.

Theorem 1.4. *Let $(y^k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$ and $(D_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{D}_\mu$ such that $\lim_{k\to\infty} D_k = \bar{D}$. If $(y^k)_{k\in\mathbb{N}}$ is quasi-Fejér monotone to a nonempty set $W \subset \mathbb{R}^n$ with respect to $(D_k)_{k\in\mathbb{N}}$ then, for each $w \in W$, the sequence $(\|y^k - w\|_{D_k})_{k\in\mathbb{N}}$ converges. Furthermore, $(y^k)_{k\in\mathbb{N}}$ is bounded and, if each cluster point of $(y^k)_{k\in\mathbb{N}}$ belongs to W, then there exists $\bar{y} \in W$ such that $\lim_{k\to\infty} y^k = \bar{y}$.*

***Proof*.** Take $w \in W$ and define the sequence $(\alpha_k)_{k\in\mathbb{N}}$, where $\alpha_k := \|y^k - w\|_{D_k}$. Since $(y^k)_{k\in\mathbb{N}}$ is quasi-Fejér monotone to W, Lemma 1.3 implies that $(\alpha_k)_{k\in\mathbb{N}}$ converges. Now, by using the first inequality in (1.3), we have $\|y^k - w\| \leq \sqrt{\mu}\alpha_k$, for all $k \in \mathbb{N}$. Thus, $(y^k)_{k\in\mathbb{N}}$ is bounded. To prove the last statement, assume that $\bar{y}, \hat{y} \in W$ are cluster points of $(y^k)_{k\in\mathbb{N}}$, and set $(y^{k_i})_{i\in\mathbb{N}}$ and $(y^{k_j})_{j\in\mathbb{N}}$ subsequences of $(y^k)_{k\in\mathbb{N}}$ such that $\lim_{j\to+\infty} y^{k_i} = \bar{y}$ and $\lim_{j\to+\infty} y^{k_j} = \hat{y}$. It follows from the first statement that $(\|y^k - \bar{y}\|_{D_k})_{k\in\mathbb{N}}$ and $(\|y^k - \hat{y}\|_{D_k})_{k\in\mathbb{N}}$ are convergent. Since $\lim_{k\to\infty} D_k = \bar{D}$, we have $\lim_{k\to\infty} \|\bar{y}\|_{D_k} = \|\bar{y}\|_{\bar{D}}$ and $\lim_{k\to\infty} \|\hat{y}\|_{D_k} = \|\hat{y}\|_{\bar{D}}$. Hence, due to

$$\langle y^k, D_k(\bar{y} - \hat{y})\rangle = \frac{1}{2}(\|y^k - \hat{y}\|^2_{D_k} - \|y^k - \bar{y}\|_{D_k} + \|\bar{y}\|_{D_k} - \|\hat{y}\|_{D_k}),$$

for all $k \in \mathbb{N}$, we conclude that the sequence $(\langle y^k, D_k(\bar{y} - \hat{y})\rangle)_{k\in\mathbb{N}}$ converges. Thus, taking into account that $\lim_{j\to+\infty} y^{k_i} = \bar{y}$, $\lim_{j\to+\infty} y^{k_j} = \hat{y}$ and $\lim_{k\to\infty} D_k = \bar{D}$ we obtain that

$$\langle \bar{y}, \bar{D}(\bar{y} - \hat{y})\rangle = \lim_{i\to\infty}\langle y^{k_i}, D_{k_i}(\bar{y} - \hat{y})\rangle = \lim_{j\to\infty}\langle y^{k_j}, D_{k_j}(\bar{y} - \hat{y})\rangle = \langle \hat{y}, \bar{D}(\bar{y} - \hat{y})\rangle.$$

Hence, using (1.3), we obtain

$$\frac{1}{\mu}\|\bar{y}-\hat{y}\|^2 \leq \|\bar{y}-\hat{y}\|_{\bar{D}}^2 = \langle \bar{y}, \bar{D}(\bar{y}-\hat{y})\rangle - \langle \hat{y}, \bar{D}(\bar{y}-\hat{y})\rangle = 0,$$

which implies that $\bar{y} = \hat{y}$. Therefore, due to $(y^k)_{k\in\mathbb{N}}$ be bounded, we conclude that $(y^k)_{k\in\mathbb{N}}$ converges to $\bar{y}$. $\square$

Chapter 2

Relative feasible inexact projections

In this chapter, we recall two concepts of relative feasible inexact projections onto a closed and convex set, and also present some new properties of them which will be used throughout this work. These concepts of inexact projections were introduced seeking to make the subproblem of computing the projections on the feasible set more efficient; see for example [14, 62, 66]. Before presenting the inexact projection concept that we will use, let us first recall the concept of exact projection with respect to a given norm. For that, *throughout this chapter* $D \in \mathcal{D}_\mu$. The *exact projection of the point* $v \in \mathbb{R}^n$ *onto* C *with respect to the norm* $\|\cdot\|_D$, denoted by $\mathcal{P}_C^D(v)$, is defined by

$$\mathcal{P}_C^D(v) := \arg\min_{z \in C} \|z - v\|_D^2. \tag{2.1}$$

?{eq:exactM}

The next result characterizes the exact projection, its proof can be found in [8, Theorem 3.14].

Lemma 2.1. *Let* $v, w \in \mathbb{R}^n$. *Then,* $w = \mathcal{P}_C^D(v)$ *if and only if* $w \in C$ *and*

$$\langle D(v - w), y - w \rangle \leq 0,$$

for all $y \in C$.

Remark 2.2. *It follows from Lemma 2.1 that* $\|\mathcal{P}_C^D(v) - \mathcal{P}_C^D(u)\|_D \leq \|v - u\|_D$. *Moreover, since* $D \in \mathcal{D}_\mu$, *by (1.3), we conclude that* $\mathcal{P}_C^D(\cdot)$ *is Lipschitz continuous with constant* $L = \mu$. *Furthermore, if* $(D_k)_{k\in\mathbb{N}} \subset \mathcal{D}_\mu$, $\lim_{k\to+\infty} z^k = \bar{z}$, *and* $\lim_{k\to+\infty} D_k = \bar{D}$, *then* $\lim_{k\to+\infty} \mathcal{P}_C^{D_k}(z^k) = \mathcal{P}_C^{\bar{D}}(\bar{z})$, *see [24, Proposition 4.2].*

In the following, we recall the concept of a feasible inexact projection with respect to $\|\cdot\|_D$ relative to a fixed point.

Definition 2.3. *The feasible inexact projection mapping, with respect to the norm* $\|\cdot\|_D$, *onto* C *relative to a point* $u \in C$ *and forcing parameter* $\zeta \in (0, 1]$, *denoted by* $\mathcal{P}_{C,\zeta}^D(u, \cdot) : \mathbb{R}^n \rightrightarrows C$, *is the set-valued mapping defined as follows*

$$\mathcal{P}_{C,\zeta}^D(u, v) := \left\{ w \in C : \ \|w - v\|_D^2 \leq \zeta \|\mathcal{P}_C^D(v) - v\|_D^2 + (1 - \zeta)\|u - v\|_D^2 \right\}. \tag{2.2}$$

?{eq:Projwm}

Each point $w \in \mathcal{P}_{C,\zeta}^D(u, v)$ *is called a feasible inexact projection, with respect to the norm* $\|\cdot\|_D$, *of* v *onto* C *relative to* u *and forcing parameter* $\zeta \in (0, 1]$.

In the following, we show that the definition given above is nothing more than a reformulation of the concept of relative feasible inexact projection with respect to $\|\cdot\|_D$ introduced in [14].

Remark 2.4. *Let $u \in C$, $v \in \mathbb{R}^n$ and D be an $n \times n$ positive definite matrix. Consider the quadratic function $Q : \mathbb{R}^n \to \mathbb{R}$ defined by $Q(z) := (1/2)\langle D(z-u), z-u\rangle + \langle D(u-v), z-u\rangle$. Thus, letting $\|\cdot\|_D$ be the norm defined by (1.2), some algebraic manipulations shows that*

$$\|z-v\|_D^2 = 2Q(z) + \|u-v\|_D^2. \tag{2.3}$$?{eq:qppq}?

Hence, (2.3) and (2.1) implies that $\mathcal{P}_C^D(v) = \arg\min_{z\in C} Q(z)$. Let $\zeta \in (0,1]$. Thus, by using (2.3), after some calculations, we can see that the following inexactness condition introduced in [14],

$$w \in C, \qquad Q(w) \le \zeta Q(\mathcal{P}_C^D(v)),$$

is equivalent to find $w \in C$ such that $\|w-v\|_D^2 \le \zeta\|\mathcal{P}_C^D(v)-v\|_D^2 + (1-\zeta)\|u-v\|_D^2$, which corresponds to condition (2.2) in Definition 2.3.

The concept of feasible inexact projection in Definition 2.3 provides more latitude to the usual concept of exact projection (2.1). The next remark makes this more precise.

Remark 2.5. *Let ζ be positive forcing parameter, $C \subset \mathbb{R}^n$ and $u \in C$ be as in Definition 2.3. First of all note that $\mathcal{P}_C^D(v) \in \mathcal{P}_{C,\zeta}^D(u,v)$. Therefore, $\mathcal{P}_{C,\zeta}^D(u,v) \neq \varnothing$, for all $u \in C$ and $v \in \mathbb{R}^n$. Consequently, the set-valued mapping $\mathcal{P}_{C,\zeta}^D(u,\cdot)$ as stated in (2.2) is well-defined. Moreover, for $\zeta = 1$, we have $\mathcal{P}_{C,1}^D(u,v) = \{\mathcal{P}_C^D(v)\}$. In addition, if $\underline{\zeta}$ and $\bar{\zeta}$ are forcing parameters such that $0 < \underline{\zeta} \le \bar{\zeta} \le 1$, then $\mathcal{P}_{C,\bar{\zeta}}^D(u,v) \subset \mathcal{P}_{C,\underline{\zeta}}^D(u,v)$.*

Lemma 2.6. *Let $v \in \mathbb{R}^n$, $u \in C$ and $w \in \mathcal{P}_{C,\zeta}^D(u,v)$. Then,*

$$\langle D(v-w), y-w\rangle \le \frac{1}{2}\|w-y\|_D^2 + \frac{1}{2}\left[\zeta\|\mathcal{P}_C^D(v)-v\|_D^2 + (1-\zeta)\|u-v\|_D^2 - \|y-v\|_D^2\right], \qquad y \in C.$$

***Proof*.** Let $y \in C$. Since $2\langle D(v-w), y-w\rangle = \|w-y\|_D^2 + \|w-v\|_D^2 - \|v-y\|_D^2$, using (2.2) we have $2\langle D(v-w), y-w\rangle = \|w-y\|_D^2 + \zeta\|\mathcal{P}_C^D(v)-v\|_D^2 + (1-\zeta)\|u-v\|_D^2 - \|v-y\|_D^2$, which is equivalent to the desired inequality. $\square$ Next, we recall a second concept of relative feasible inexact projection onto a closed convex set, see [2, 29]. The definition is as follows.

Definition 2.7. *The feasible inexact projection mapping, with respect to the norm $\|\cdot\|_D$, onto C relative to $u \in C$ and forcing parameter $\gamma \ge 0$, denoted by $\mathcal{R}_{C,\gamma}^D(u,\cdot) : \mathbb{R}^n \rightrightarrows C$, is the set-valued mapping defined as follows*

$$\mathcal{R}_{C,\gamma}^D(u,v) := \left\{w \in C : \ \langle D(v-w), y-w\rangle \le \gamma\|w-u\|_D^2, \quad \forall\, y \in C\right\}. \tag{2.4}$$?{eq:Projw}?

Each point $w \in \mathcal{R}_{C,\gamma}^D(u,v)$ is called a feasible inexact projection, with respect to the norm $\|\cdot\|_D$, of v onto C relative to u and forcing parameter $\gamma \ge 0$.

The concept of feasible inexact projection in Definition 2.7 also provides more latitude to the usual concept of exact projection. Next, we present some remarks about this concept.

Remark 2.8. *Let $\gamma \geq 0$ be a forcing parameter, $C \subset \mathbb{R}^n$ and $u \in C$ be as in Definition 2.7. For all $v \in \mathbb{R}^n$, it follows from (2.4) and Lemma 2.1 that $\mathcal{R}^D_{C,0}(u,v) = \{\mathcal{P}^D_C(v)\}$ is the exact projection of v onto C. Moreover, $\mathcal{P}^D_C(v) \in \mathcal{R}^D_{C,\gamma}(u,v)$ concluding that $\mathcal{R}_{C,\gamma}(u,v) \neq \varnothing$, for all $u \in C$ and $v \in \mathbb{R}^n$. Consequently, the set-valued mapping $\mathcal{R}^D_{C,\gamma}(u,\cdot)$ as stated in (2.4) is well-defined.*

The next lemma is a variation of [31, Lemma 6]. It will allow to relate Definitions 2.3 and 2.7.

Lemma 2.9. *Let $v \in \mathbb{R}^n$, $u \in C$, $\gamma \geq 0$ and $w \in \mathcal{R}^D_{C,\gamma}(u,v)$. Then, for all $0 \leq \gamma < 1/2$, the holds*

$$\|w - x\|^2_D \leq \|x - v\|^2_D + \frac{2\gamma}{1-2\gamma}\|u - v\|^2_D - \frac{1}{1-2\gamma}\|w - v\|^2_D, \qquad \forall x \in C.$$

Proof. First note that $\|w - x\|^2_D = \|x - v\|^2_D - \|w - v\|^2_D + 2\langle D(v - w), x - w\rangle$. Since $w \in \mathcal{R}^D_{C,\gamma}(u,v)$ and $x \in C$, combining the last equality with (2.4), we obtain

$$\|w - x\|^2_D \leq \|x - v\|^2_D - \|w - v\|^2_D + 2\gamma\|w - u\|^2_D. \tag{2.5}$$?{eq:fg}?

On the other hand, we also have $\|w - u\|^2_D = \|u - v\|^2_D - \|w - v\|^2_D + 2\langle D(v - w), u - w\rangle$. Due to $w \in \mathcal{R}^D_{C,\gamma}(u,v)$ and $u \in C$, using (2.4) and considering that $0 \leq \gamma < 1/2$, we have

$$\|w - u\|^2_D \leq \frac{1}{1-2\gamma}\|u - v\|^2_D - \frac{1}{1-2\gamma}\|w - v\|^2_D.$$

Therefore, substituting the last inequality into (2.5), we obtain the desired inequality. □

In the following lemma, we present a relationship between Definitions 2.3 and 2.7.

Lemma 2.10. *Let $v \in \mathbb{R}^n$, $u \in C$, $\gamma \geq 0$ and $\zeta \in (0,1]$. If $0 \leq \gamma < 1/2$ and $\zeta = 1 - 2\gamma$, then*

$$\mathcal{R}^D_{C,\gamma}(u,v) \subset \mathcal{P}^D_{C,\zeta}(u,v).$$

Proof. Let $w \in \mathcal{R}^D_{C,\gamma}(u,v)$. Applying Lemma 2.9 with $x = \mathcal{P}^D_C(v)$ we have

$$\|w - \mathcal{P}^D_C(v)\|^2_D \leq \|v - \mathcal{P}^D_C(v)\|^2_D + \frac{2\gamma}{1-2\gamma}\|u - v\|^2_D - \frac{1}{1-2\gamma}\|w - v\|^2_D,$$

After some algebraic manipulations in the last inequality we obtain that

$$\|w - v\|^2_D \leq (1-2\gamma)\|v - \mathcal{P}^D_C(v)\|^2_D + 2\gamma\|u - v\|^2_D - (1-2\gamma)\|w - \mathcal{P}^D_C(v)\|^2_D.$$

Therefore, considering that $0 \leq \gamma < 1/2$ and $\zeta = 1-2\gamma$, the result follows from Definition 2.3. □

Remark 2.11. *Under the conditions of Lemma 2.10, there exists $0 \le \gamma < 1/2$ and $\zeta = 1-2\gamma$ such that $\mathcal{P}^D_{C,\zeta}(u,v) \nsubseteq \mathcal{R}^D_{C,\gamma}(u,v)$. Indeed, let $\gamma = 3/8$, $\zeta = 1/4$, and $\bar{w} = \frac{1}{2}(\mathcal{P}^D_C(v)+u)$, then*

$$\|\bar{w}-v\|_D^2 = \frac{1}{4}\|\mathcal{P}^D_C(v)-v\|_D^2 + \frac{1}{4}\|u-v\|_D^2 + \frac{1}{2}\langle D(\mathcal{P}^D_C(v)-v), u-v\rangle.$$

Since $\mathcal{P}^D_C(v)$ is the exact projection of v, we have $\langle D(\mathcal{P}^D_C(v)-v), u-v\rangle \le \|u-v\|_D^2$. Combining this inequality with the last equality and Definition 2.3, we conclude that $\bar{w} \in \mathcal{P}^D_{C,\zeta}(u,v)$. Now, letting $w_t = t\mathcal{P}^D_C(v) + (1-t)\bar{w}$ with $0 < t < 1$, after some algebraic manipulations we have

$$\langle D(v-\bar{w}), w_t-\bar{w}\rangle = t\|\bar{w}-u\|_D^2 - \frac{t}{2}\langle D(v-\mathcal{P}^D_C(v)), u-\mathcal{P}^D_C(v)\rangle.$$

Thus, it follows from Lemma 2.1 that $\langle D(v-\bar{w}), w_t-\bar{w}\rangle \ge t\|\bar{w}-u\|_D^2$. Hence, taking $t > 3/8$ we conclude that $\bar{w} \notin \mathcal{R}^D_{C,\gamma}(u,v)$. Therefore, considering that $\bar{w} \in \mathcal{P}^D_{C,\zeta}(u,v)$, the statement follows.

It follows from Remark 2.11 that, in general, $\mathcal{P}^D_{C,\zeta}(u,v) \nsubseteq \mathcal{R}^D_{C,\gamma}(u,v)$. However, whenever C is a bounded set, we will show that for each fixed $0 \le \gamma < 1/2$ there exist $0 < \zeta < 1$ such that $\mathcal{P}^D_{C,\zeta}(u,v) \subseteq \mathcal{R}^D_{C,\gamma}(u,v)$. For that, we first need the next lemma.

Lemma 2.12. *Let $v \in \mathbb{R}^n$, $u \in C$ and $0 < \gamma < 1/2$. Assume that C is a bounded set and take*

$$0 < \varepsilon < \frac{\gamma\|u-\mathcal{P}^D_C(v)\|_D^2}{1-\gamma+\|v-\mathcal{P}^D_C(v)\|_D + 2\gamma\|u-\mathcal{P}^D_C(v)\|_D + \operatorname{diam}C}, \tag{2.6}$$?{eq:epsi}?

where diam$C$ *denotes the diameter of C. Then, $\{w \in C : \|w-\mathcal{P}^D_C(v)\|_D \le \varepsilon\} \subset \mathcal{R}^D_{C,\gamma}(u,v)\}$.*

***Proof*.** Take ε satisfying (2.6) and $w \in C$ such that $\|w-\mathcal{P}^D_C(v)\|_D \le \varepsilon$. For all $z \in C$, we have

$$\begin{aligned}\langle D(v-w), z-w\rangle = \langle D(v-\mathcal{P}^D_C(v)), z-\mathcal{P}^D_C(v)\rangle + \langle D(v-\mathcal{P}^D_C(v)), \mathcal{P}^D_C(v)-w\rangle \\ + \langle D(\mathcal{P}^D_C(v)-w), z-\mathcal{P}^D_C(v)\rangle + \|\mathcal{P}^D_C(v)-w\|_D^2.\end{aligned}$$

Using Lemma 2.1, we have $\langle D(v-\mathcal{P}^D_C(v)), z-\mathcal{P}^D_C(v)\rangle \le 0$. Thus, the last equality becomes

$$\langle D(v-w), z-w\rangle \le \langle D(v-\mathcal{P}^D_C(v)), \mathcal{P}^D_C(v)-w\rangle + \langle D(\mathcal{P}^D_C(v)-w), z-\mathcal{P}^D_C(v)\rangle + \|\mathcal{P}^D_C(v)-w\|_D^2.$$

By using Cauchy-Schwarz inequality, we conclude from the last inequality that

$$\langle D(v-w), z-w\rangle \le \|w-\mathcal{P}^D_C(v)\|_D \left(\|v-\mathcal{P}^D_C(v)\|_D + \|z-\mathcal{P}^D_C(v)\|_D\right) + \|w-\mathcal{P}^D_C(v)\|_D^2.$$

Since $\|w-\mathcal{P}^D_C(v)|_D \le \varepsilon$ and $\|z-\mathcal{P}^D_C(v)\|_D \le \operatorname{diam}C$, the last inequality implies that

$$\langle D(v-w), z-w\rangle \le \varepsilon\left(\|v-\mathcal{P}^D_C(v)\|_D + \operatorname{diam}C\right) + \varepsilon^2, \tag{2.7}$$?{eq:diam1}?

On the other hand, if ε satisfies (2.6) then

$$\varepsilon\left(1-\gamma+\|v-\mathcal{P}^D_C(v)\|_D + \operatorname{diam}C\right) + \gamma\varepsilon^2 < \gamma\|u-\mathcal{P}^D_C(v)\|_D^2 - 2\gamma\varepsilon\|u-\mathcal{P}^D_C(v)\|_D + \gamma\varepsilon^2,$$

hence $\varepsilon\left(1-\gamma+\|v-\mathcal{P}_C^D(v)\|_D+\operatorname{diam}C\right)+\gamma\varepsilon^2<\gamma\left(\|u-\mathcal{P}_C^D(v)\|_D-\varepsilon\right)^2$. Since $\gamma,\varepsilon<1$, we have $\varepsilon^2<\varepsilon(1-\gamma)+\gamma\varepsilon^2$ and we can conclude that

$$\varepsilon\left(\|v-\mathcal{P}_C^D(v)\|_D+\operatorname{diam}C\right)+\varepsilon^2<\gamma\left(\|u-\mathcal{P}_C^D(v)\|_D-\varepsilon\right)^2.$$

It follows from (2.7) that

$$\langle D(v-w),z-w\rangle\leq\gamma\left(\|u-\mathcal{P}_C^D(v)\|_D-\varepsilon\right)^2. \tag{2.8}$$?{eq:diam2}?

Using again that $\|w-\mathcal{P}_C^D(v)|_D\leq\varepsilon$ and the triangular inequality, we have

$$0<\|u-\mathcal{P}_C^D(v)\|_D-\varepsilon\leq\|u-\mathcal{P}_C^D(v)\|_D-\|w-\mathcal{P}_C^D(v)\|_D\leq\|u-w\|_D.$$

Hence, taking into account (2.8), we conclude that $\langle D(v-w),z-w\rangle\leq\gamma\|u-w\|_D^2$. Therefore, it follows from Definition 2.7 that $w\in\mathcal{R}_{C,\gamma}^D(u,v)$. □

Proposition 2.13. *Let $v\in\mathbb{R}^n$, $u\in C$ and assume that C is a bounded set. Then, for each $0<\gamma<1/2$, there exist $0<\zeta<1$ such that $\mathcal{P}_{C,\zeta}^D(u,v)\subseteq\mathcal{R}_{C,\gamma}^D(u,v)$.*

Proof. It follows from Lemma 2.12 that given $0<\gamma<1/2$ there exists $\varepsilon>0$ such that, for all $w\in C$ with $\|w-\mathcal{P}_C^D(v)\|\leq\varepsilon$, we have $w\in\mathcal{R}_\gamma^D(v)$. Otherwise, we can see in (2.2), when $\zeta\to1$, the diameter of $C\cap\mathcal{P}_{C,\zeta}^D(u,v)$ tends to zero, then there exists ζ close to 1 such that $\operatorname{diam}(C\cap\mathcal{P}_{C,\zeta}^D(u,v))<\varepsilon/2$, and $\mathcal{P}_{C,\zeta}^D(u,v)\subset\mathcal{R}_{C,\gamma}^D(u,v)$. □ Next, we present important properties of inexact projections, it will be useful in the sequel.

Lemma 2.14. *Let $x\in C$, $\alpha>0$ and $z(\alpha)=x-\alpha D^{-1}\nabla f(x)$. Take $w(\alpha)\in\mathcal{P}_{C,\zeta}^D(x,z(\alpha))$ with $\zeta\in(0,1]$. Then,*

(i) $\langle\nabla f(x),w(\alpha)-x\rangle\leq-\dfrac{1}{2\alpha}\|w(\alpha)-x\|_D^2+\dfrac{\zeta}{2\alpha}\left[\|\mathcal{P}_C^D(z(\alpha))-z(\alpha)\|_D^2-\|x-z(\alpha)\|_D^2\right]$;

(ii) the point x is stationary for problem (1) if and only if $x\in\mathcal{P}_{C,\zeta}^D(x,z(\alpha))$;

(iii) if $x\in C$ is a nonstationary point for problem (1), then $\Big\langle\nabla f(x),w(\alpha)-x\Big\rangle<0$. Equivalently, if there exists $\bar\alpha>0$ such that $\Big\langle\nabla f(x),w(\bar\alpha)-x\Big\rangle\geq0$, then x is stationary for problem (1).

Proof. Since $w(\alpha)\in\mathcal{P}_{C,\zeta}^D(x,z(\alpha))$, applying Lemma 2.6 with $w=w(\alpha)$, $v=z(\alpha)$, $y=x$, and $u=x$, we conclude, after some algebraic manipulations, that

$$\langle D(z(\alpha)-w(\alpha)),x-w(\alpha)\rangle\leq\frac{1}{2}\|w(\alpha)-x\|_D^2+\frac{\zeta}{2}\left[\|\mathcal{P}_C^D(z(\alpha))-z(\alpha)\|_D^2-\|x-z(\alpha)\|_D^2\right].$$

Substituting $z(\alpha)=x-\alpha\nabla f(x)$ in the left hand side of the last inequality, some manipulations yield the inequality of item (i). For proving item (ii), we first assume that x is stationary for problem (1). In this case, (1.1) implies that $\langle\nabla f(x),w(\alpha)-x\rangle\geq0$. Hence, due to $\|\mathcal{P}_C^D(z(\alpha))-z(\alpha)\|_D\leq\|x-z(\alpha)\|_D$, item (i) implies

$$\frac{1}{2\alpha}\|w(\alpha)-x\|_D^2\leq\frac{\zeta}{2\alpha}\left[\|\mathcal{P}_C^D(z(\alpha))-z(\alpha)\|_D^2-\|x-z(\alpha)\|_D^2\right]\leq0.$$

Since $\alpha > 0$ and $\zeta \in (0,1]$, the last inequality yields $w(\alpha) = x$. Therefore, $x \in \mathcal{P}^D_{C,\zeta}(x, z(\alpha))$. Reciprocally, if $x \in \mathcal{P}^D_{C,\zeta}(x, z(\alpha))$, then Definition 2.3 implies that

$$\|x - z(\alpha)\|_D^2 \leq \zeta \|\mathcal{P}^D_C(z(\alpha)) - z(\alpha)\|_D^2 + (1-\zeta)\|x - z(\alpha)\|_D^2.$$

Hence, $0 \leq \zeta \left(\|\mathcal{P}^D_C(z(\alpha)) - z(\alpha)\|_D^2 - (\|x - z(\alpha)\|_D^2 \right)$. Considering that $\zeta \in (0,1]$ we have

$$\|x - z(\alpha)\|_D \leq \|\mathcal{P}^D_C(z(\alpha)) - z(\alpha)\|_D.$$

Thus, due to exact projection with respect to the norm $\|\cdot\|_D$ be unique and $z(\alpha) = x - D^{-1}\alpha\nabla f(x)$, we have $\mathcal{P}^D_C(x - \alpha D^{-1}\nabla f(x)) = x$. Hence, x is the solution of the constrained optimization problem $\min_{y\in C} \|y - z(\alpha)\|_D^2$, which taking into account that $\alpha > 0$ implies (1.1). Therefore, x is stationary point for problem (1). Finally, to prove item (iii), take x a nonstationary point for problem (1). Thus, by item (ii), $x \notin \mathcal{P}^D_{C,\zeta}(x, z(\alpha))$ and taking into account that $w(\alpha) \in \mathcal{P}^D_{C,\zeta}(x, z(\alpha))$, we conclude that $x \neq w(\alpha)$. Since $\|\mathcal{P}^D_C(z(\alpha)) - z(\alpha)\|_D \leq \|x - z(\alpha)\|_D$, $\alpha > 0$ and $\zeta \in (0,1]$, it follows from item (i) that $\left\langle \nabla f(x), w(\alpha) - x \right\rangle < 0$ and the first sentence is proved. Finally, note that the second sentence is the contrapositive of the first sentence. $\square$ Finally, it is worth mentioning that Definitions 2.3 and 2.7, introduced respectively in [14] and [29], are relative inexact concepts, while the concept introduced in [62, 66] is absolute.

2.1 Practical computation of inexact projections

In this section, for a given $v \in \mathbb{R}^n$ and $u \in C$, we discuss how to calculate a point $w \in C$ belonging to $\mathcal{P}^D_{C,\zeta}(u, v)$ or $\mathcal{R}^D_{C,\gamma}(u, v)$. We recall that Lemma 2.10 implies that $\mathcal{P}^D_{C,\zeta}(u, v)$ has more latitude than $\mathcal{R}^D_{C,\gamma}(u, v)$, i.e., $\mathcal{R}^D_{C,\gamma}(u, v) \subset \mathcal{P}^D_{C,\zeta}(u, v)$.

We begin our discussion by showing how a point $w \in \mathcal{P}^D_{C,\zeta}(u, v)$ can be calculated without knowing the point $\mathcal{P}^D_C(v)$. Considering that this discussion has already been covered in [14, Section 3, Algorithm 3.1], we will limit ourselves to giving a general idea of how this task is carried out; see also [17, Section 5.1]. The idea is to use an external procedure capable of computing two sequences $(c_\ell)_{\ell\in\mathbb{N}} \subset \mathbb{R}$ and $(w^\ell)_{\ell\in\mathbb{N}} \subset C$ satisfying the following conditions

$$c_\ell \leq \|\mathcal{P}^D_C(v) - v\|_D^2, \quad \forall \ell \in \mathbb{N}, \qquad \lim_{\ell\to+\infty} c_\ell = \|\mathcal{P}^D_C(v) - v\|_D^2, \qquad \lim_{\ell\to+\infty} w^\ell = \mathcal{P}^D_C(v). \tag{2.9}$$

?{def:cl}?

In this case, if $v \notin C$, then we have $\|\mathcal{P}^D_C(v) - v\|_D^2 - \|u - v\|_D^2 < 0$. Hence, given an arbitrary $\zeta \in (0,1)$, the second condition in (2.9) implies that there exists $\hat{\ell}$ such that

$$\|\mathcal{P}^D_C(v) - v\|_D^2 - \|u - v\|_D^2 < \zeta(c_{\hat{\ell}} - \|u - v\|_D^2).$$

Moreover, by using the last condition in (2.9), we conclude that there exists $\bar{\ell} > \hat{\ell}$ such that

$$\|w_{\bar{\ell}} - v\|_D^2 - \|u - v\|_D^2 < \zeta(c_{\bar{\ell}} - \|u - v\|_D^2), \tag{2.10}$$

?{def:clsc}?

which using the inequality in (2.9) yields $\|w_{\bar{\ell}} - v\|_D^2 < \zeta \|\mathcal{P}_C^D(v) - v\|_D^2 + (1-\zeta)\|u - v\|_D^2$. Hence, Definition 2.3 implies that $w_{\bar{\ell}} \in \mathcal{P}_{C,\zeta}^D(u,v)$. Therefore, (2.10) can be used as a stopping criterion to compute a feasible inexact projection, with respect to the norm $\|\cdot\|_D$, of v onto C relative to u and forcing parameter $\zeta \in (0,1]$. For instance, it follows from [14, Theorem 3.2, Lemma 3.1] (see also [16]) that such sequences $(c_\ell)_{\ell\in\mathbb{N}} \subset \mathbb{R}$ and $(w^\ell)_{\ell\in\mathbb{N}} \subset C$ satisfying (2.9) can be computed by using *Dykstra's algorithm* [23, 33], whenever D is the identity matrix and the set $C = \cap_{i=1}^p C_i$, where C_i are closed and convex sets and the exact projection $\mathcal{P}_{C_i}^D(v)$ is easy to obtain, for all $i = 1, \ldots, p$.

We end this section by discussing how to compute a point $w \in \mathcal{R}_{C,\gamma}^D(u,v)$. For that, we apply the classical *Frank-Wolfe method*, also known as *conditional gradient method*, to minimize the function $\psi(z) := \|z - v\|^2/2$ onto the constraint set C with a suitable stop criteria depending of $u \in C$ and $\gamma \in (0,1]$, see [9, 49]. To state the method we assume the existence of a linear optimization oracle (or simply LO oracle) capable of minimizing linear functions over the constraint set C, which is assumed to be compact. The Frank-Wolfe method is formally stated as follows.

Algorithm 2.1 Frank-Wolfe method to compute $w \in \mathcal{R}_{C,\gamma}^D(u,v)$

Input: $D \in \mathcal{D}_\mu$, $\gamma \in (0,1]$, $v \in \mathbb{R}^n$ and $u \in C$.

Step 0. Let $w^0 \in C$ and set $\ell \leftarrow 0$.

Step 1. Use a LO oracle to compute an optimal solution z^ℓ and the optimal value s_ℓ^* as

$$z^\ell \in \arg\min_{z\in C} \langle w^\ell - v,\ z - w^\ell\rangle, \qquad s_\ell^* := \langle w^\ell - v,\ z^\ell - w^\ell\rangle. \tag{2.11}$$?{eq:CondG_{

If $-s_\ell^* \le \gamma\|w^\ell - u\|_D^2$, then define $w := w^\ell$ and **stop**.

Step 2. Compute α_ℓ and $w_{\ell+1}$ as

$$w_{\ell+1} := w^\ell + \alpha_\ell(z^\ell - w^\ell), \qquad \alpha_\ell := \min\left\{1, -s_\ell^*/\|z^\ell - w^\ell\|^2\right\}. \tag{2.12}$$?{eq:step si

Set $\ell \leftarrow \ell + 1$, and go to Step 1.

Output: $w := w^\ell$.

Let us describe the main features of Algorithm 2.1, i.e., the Frank-Wolfe method applied to the problem $\min_{z\in C}\psi(z)$. In this case, (2.11) is equivalent to $s_\ell^* := \min_{z\in C}\langle \psi'(w^\ell),\ z - w^\ell\rangle$. Since ψ is convex, we have $\psi(z) \ge \psi(w^\ell) + \langle \psi'(w^\ell),\ z - w^\ell\rangle \ge \psi(w^\ell) + s_\ell^*$, for all $z \in C$. Define $w_* := \arg\min_{z\in C}\psi(z)$ and $\psi^* := \min_{z\in C}\psi(z)$. Letting $z = w_*$ in the last inequality, we obtain $\psi(w^\ell) \ge \psi^* \ge \psi(w^\ell) + s_\ell^*$, which implies that $s_\ell^* < 0$ whenever $\psi(w^\ell) \ne \psi^*$. Thus, we conclude that $-s_\ell^* = \langle v - w^\ell,\ z^\ell - w^\ell\rangle > 0 \ge \langle v - w_*,\ z - w_*\rangle$, for all $z \in C$. Therefore, if Algorithm 2.1 computes $w^\ell \in C$ satisfying $-s_\ell^* \le \gamma\|w^\ell - u\|_D^2$, then the method terminates. Otherwise, it computes the step size $\alpha_\ell = \arg\min_{\alpha\in[0,1]}\psi(w^\ell + \alpha(z^\ell - w^\ell))$ using exact minimization. Since z^ℓ, $w^\ell \in C$ and C is convex, we conclude from (2.12) that $w_{\ell+1} \in C$, thus Algorithm 2.1 generates a sequence in C. Finally, (2.11) implies that $\langle v - w^\ell,\ z - w^\ell \rangle \leq -s^*_\ell$, for all $z \in C$. Considering that [9, Proposition A.2] implies that $\lim_{\ell \to +\infty} s^*_\ell = 0$ and taking into account the stopping criteria $-s^*_\ell \leq \gamma \|w^\ell - u\|^2_D$, we conclude that the output of Algorithm 2.1 is a feasible inexact projection $w \in \mathcal{R}^D_{C,\gamma}(u, v)$ i.e., $\langle v - w,\ z - w \rangle \leq \gamma \|w^\ell - u\|^2_D$, for all $z \in C$.

Bibliography

r2021inexact? [1] A. A. Aguiar, O. P. Ferreira, and L. F. Prudente. Inexact gradient projection method with relative error tolerance. *arXiv preprint arXiv:2101.11146*, 2021.

n_Leandro2020 [2] A. A. Aguiar, O. P. Ferreira, and L. F. Prudente. Subgradient method with feasible inexact projections for constrained convex optimization problems. *Optimization*, 0(0):1–23, 2021, https://doi.org/10.1080/02331934.2021.1902520.

Ahookhosh2012 [3] M. Ahookhosh, K. Amini, and S. Bahrami. A class of nonmonotone Armijo-type line search method for unconstrained optimization. *Optimization*, 61(4):387–404, 2012.

en2017linear? [4] Z. Allen-Zhu, E. Hazan, W. Hu, and Y. Li. Linear convergence of a Frank-Wolfe type algorithm over trace-norm balls. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 6192–6201, Red Hook, NY, USA, 2017. Curran Associates Inc.

imanum/drh020 [5] R. Andreani, E. G. Birgin, J. M. Martínez, and J. Yuan. Spectral projected gradient and variable metric methods for optimization with linear inequalities. *IMA J. Numer. Anal.*, 25(2):221–252, 04 2005, https://academic.oup.com/imajna/article-pdf/25/2/221/2090233/drh020.pdf.

?Paulo2007? [6] A. Auslender, P. J. S. Silva, and M. Teboulle. Nonmonotone projected gradient methods based on barrier and Euclidean distances. *Comput. Optim. Appl.*, 38(3):305–327, 2007.

BB1988 [7] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8(1):141–148, 1988.

chkeLivro2014 [8] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer Publishing Company, Incorporated, 1st edition, 2011.

kTeboulle2004 [9] A. Beck and M. Teboulle. A conditional gradient method with linear rate of convergence for solving convex linear systems. *Math. Methods Oper. Res.*, 59(2):235–247, 2004.

er_roman2010 [10] J. Y. Bello Cruz and L. R. Lucambio Pérez. Convergence of a projected gradient method variant for quasiconvex objectives. *Nonlinear Anal.*, 73(9):2917–2922, 2010.

ertsekas1976 [11] D. P. Bertsekas. On the Goldstein-Levitin-Polyak gradient projection method. *IEEE Trans. Automat. Control*, 21(2):174–184, 1976.

ertsekas1999 [12] D. P. Bertsekas. *Nonlinear programming.* Athena Scientific Optimization and Computation Series. Athena Scientific, Belmont, MA, second edition, 1999.

spgsiam [13] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, 10(4):1196–1211, 2000, https://doi.org/10.1137/S1052623497330963.

ezRaydan2003 [14] E. G. Birgin, J. M. Martínez, and M. Raydan. Inexact spectral projected gradient methods on convex sets. *IMA J. Numer. Anal.*, 23(4):539–559, 2003.

?JSSv060i03? [15] E. G. Birgin, J. M. Martínez, and M. Raydan. Spectral projected gradient methods: Review and perspectives. *J. Stat. Softw.*, 60(3):1–21, 2014.

nRaydan2005 [16] E. G. Birgin and M. Raydan. Robust stopping criteria for dykstra's algorithm. *SIAM J. Sci. Comput.*, 26(4):1405–1414, 2005.

onettini2016 [17] S. Bonettini, I. Loris, F. Porta, and M. Prato. Variable metric inexact line-search-based methods for nonsmooth optimization. *SIAM J. Optim.*, 26(2):891–921, 2016.

i2019recent [18] S. Bonettini, F. Porta, M. Prato, S. Rebegoldi, V. Ruggiero, and L. Zanni. Recent advances in variable metric first-order methods. In *Computational Methods for Inverse Problems in Imaging*, pages 1–31. Springer, 2019.

niPrato2015 [19] S. Bonettini and M. Prato. New convergence results for the scaled gradient projection method. *Inverse Problems*, 31(9):095008, 20, 2015.

onettini2009 [20] S. Bonettini, R. Zanella, and L. Zanni. A scaled gradient projection method for constrained image deblurring. *Inverse Problems*, 25(1):015002, 23, 2009.

Nocedal2018 [21] L. Bottou, F. E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM Rev.*, 60(2):223–311, 2018.

_Ferron1994 [22] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* Society for Industrial and Applied Mathematics, 1994, https://epubs.siam.org/doi/pdf/10.1137/1.9781611970777.

Dykstra1986 [23] J. P. Boyle and R. L. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. In *Advances in order restricted statistical inference (Iowa City, Iowa, 1985)*, volume 37 of *Lect. Notes Stat.*, pages 28–47. Springer, Berlin, 1986.

ettesVu2013 [24] P. L. Combettes and B. C. Vũ. Variable metric quasi-Fejér monotonicity. *Nonlinear Anal.*, 78:17–31, 2013.

Dai2002 [25] Y. H. Dai. On the nonmonotone line search. *J. Optim. Theory Appl.*, 112(2):315–330, 2002.

Fletcher2005 [26] Y.-H. Dai and R. Fletcher. Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming. *Numer. Math.*, 100(1):21–47, 2005.

Fletcher2006 [27] Y.-H. Dai and R. Fletcher. New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. *Math. Program.*, 106(3, Ser. A):403–421, 2006.

DaiHage2006 [28] Y.-H. Dai, W. W. Hager, K. Schittkowski, and H. Zhang. The cyclic Barzilai-Borwein method for unconstrained optimization. *IMA J. Numer. Anal.*, 26(3):604–627, 2006.

aGilson2018 [29] F. R. de Oliveira, O. P. Ferreira, and G. N. Silva. Newton's method with feasible inexact projections for solving constrained generalized equations. *Comput. Optim. Appl.*, 72(1):159–177, 2019.

Serafino2018 [30] D. di Serafino, V. Ruggiero, G. Toraldo, and L. Zanni. On the steplength selection in gradient methods for unconstrained optimization. *Appl. Math. Comput.*, 318:176–195, 2018.

Leandro2019 [31] R. Díaz Millán, O. P. Ferreira, and L. F. Prudente. Alternating conditional gradient method for convex feasibility problems. *arXiv e-prints*, page arXiv:1912.04247, Dec 2019, 1912.04247.

enchmarking? [32] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2):201–213, 2002.

Dykstra1983 [33] R. L. Dykstra. An algorithm for restricted least squares regression. *J. Amer. Statist. Assoc.*, 78(384):837–842, 1983.

Wang_Yan2019 [34] J. Fan, L. Wang, and A. Yan. An inexact projected gradient method for sparsity-constrained quadratic measurements regression. *Asia-Pac. J. Oper. Res.*, 36(2):1940008, 21, 2019.

huverdt2019? [35] N. S. Fazzio and M. L. Schuverdt. Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems. *Optim. Lett.*, 13(6):1365–1379, 2019.

gueiredo2007 [36] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Topics Signal Process.*, 1(4):586–597, Dec 2007.

edlander1999 [37] A. Friedlander, J. M. Martínez, B. Molina, and M. Raydan. Gradient method with retards and generalizations. *SIAM J. Numer. Anal.*, 36(1):275–289, 1999.

erHazan2015? [38] D. Garber and E. Hazan. Faster rates for the frank-wolfe method over strongly-convex sets. *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, pages 541–549, 2015.

e_Davies2018 [39] M. Golbabaee and M. E. Davies. Inexact gradient projection and fast data driven compressed sensing. *IEEE Trans. Inform. Theory*, 64(10):6707–6721, 2018.

oldstein1964 [40] A. A. Goldstein. Convex programming in Hilbert space. *Bull. Amer. Math. Soc.*, 70:709–710, 1964.

asprojected? [41] D. S. Gonçalves, M. A. Gomes-Ruggiero, and C. Lavor. A projected gradient method for optimization over density matrices. *Optim. Methods Softw.*, 31(2):328–341, 2016, https://doi.org/10.1080/10556788.2015.1082105.

ncalves2020 [42] D. S. Gonçalves, M. L. N. Gonçalves, and T. C. Menezes. Inexact variable metric method for convex-constrained optimization problems. *Optimization*, 0(0):1–19, 2021, https://doi.org/10.1080/02331934.2021.1887181.

Gong2011 [43] P. Gong, K. Gai, and C. Zhang. Efficient euclidean projections via piecewise root finding and its application in gradient projection. *Neurocomputing*, 74(17):2754 – 2766, 2011.

iaSachs2017 [44] G. N. Grapiglia and E. W. Sachs. On the worst-case evaluation complexity of non-monotone line search algorithms. *Comput. Optim. Appl.*, 68(3):555–577, 2017.

iaSachs2020 [45] G. N. Grapiglia and E. W. Sachs. A generalized worst-case complexity analysis for non-monotone line searches. *Numer. Algorithms*, 87(2):779–796, Jun 2021.

Grippo1986 [46] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, 23(4):707–716, 1986.

Higham2002 [47] N. J. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal.*, 22(3):329–343, 2002.

Iusem2003 [48] A. N. Iusem. On the convergence properties of the projected gradient method for convex optimization. *Comput. Appl. Math.*, 22(1):37–52, 2003.

Jaggi2013 [49] M. Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In S. Dasgupta and D. McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 427–435, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

_Levitin1966 [50] E. Levitin and B. Polyak. Constrained minimization methods. *USSR Comput. Math. Math. Phys.*, 6(5):1 – 50, 1966.

_Hu_Gao2015 [51] G. Ma, Y. Hu, and H. Gao. An accelerated momentum based gradient projection method for image deblurring. In *2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–4, 2015.

MoLiuYan2007 [52] J. Mo, C. Liu, and S. Yan. A nonmonotone trust region method based on nonincreasing technique of weighted average of the successive function values. *J. Comput. Appl. Math.*, 209(1):97–108, 2007.

More1990 [53] J. J. Moré. On the performance of algorithms for large-scale bound constrained problems. In *Large-scale numerical optimization (Ithaca, NY, 1989)*, pages 32–45. SIAM, Philadelphia, PA, 1990.

irovski2013 [54] Y. Nesterov and A. Nemirovski. On first-order algorithms for ℓ_1/nuclear norm minimization. *Acta Numer.*, 22:509–575, 2013.

06numerical? [55] J. Nocedal and S. Wright. *Numerical optimization.* Springer Science & Business Media, 2006.

Panier1991 [56] E. R. Panier and A. L. Tits. Avoiding the Maratos effect by means of a nonmonotone line search. I. General constrained problems. *SIAM J. Numer. Anal.*, 28(4):1183–1195, 1991.

Necoara2018 [57] A. Patrascu and I. Necoara. On the convergence of inexact projection primal first-order methods for convex minimization. *IEEE Trans. Automat. Control*, 63(10):3317–3329, 2018.

akLivro1987 [58] B. T. Polyak. *Introduction to optimization.* Translations Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York, 1987. Translated from the Russian, With a foreword by Dimitri P. Bertsekas.

Rasch2020 [59] J. Rasch and A. Chambolle. Inexact first-order primal-dual algorithms. *Comput. Optim. Appl.*, 76(2):381–430, 2020.

?dykstraSDD? [60] M. Raydan and P. Tarazaga. Primal and polar approach for computing the symmetric diagonally dominant projection. *Numer. Linear Algebra Appl.*, 9(5):333–345, 2002, https://onlinelibrary.wiley.com/doi/pdf/10.1002/nla.277.

hsSachs2011 [61] E. W. Sachs and S. M. Sachs. Nonmonotone line searches for optimization algorithms. *Control Cybernet.*, 40(4):1059–1075, 2011.

zoVilla2012 [62] S. Salzo and S. Villa. Inexact and accelerated proximal point algorithms. *J. Convex Anal.*, 19(4):1167–1192, 2012.

_Wright2012 [63] S. Sra, S. Nowozin, and S. Wright. *Optimization for Machine Learning.* Neural information processing series. MIT Press, 2012.

_davies2017 [64] J. Tang, M. Golbabaee, and M. E. Davies. Gradient projection iterative sketch for large-scale constrained least-squares. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3377–3386, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

Toint1996 [65] P. L. Toint. An assessment of nonmonotone linesearch techniques for unconstrained optimization. *SIAM J. Sci. Comput.*, 17(3):725–739, 1996.

llaSalzo2013 [66] S. Villa, S. Salzo, L. Baldassarre, and A. Verri. Accelerated and inexact forward-backward algorithms. *SIAM J. Optim.*, 23(3):1607–1633, 2013.

WangLiu2005? [67] C. Wang, Q. Liu, and X. Yang. Convergence properties of nonmonotone spectral projected gradient methods. *J. Comput. Appl. Math.*, 182(1):51–66, 2005.

Xihong2018 [68] X. Yan, K. Wang, and H. He. On the convergence rate of scaled gradient projection method. *Optimization*, 67(9):1365–1376, 2018.

ng_Yang2019 [69] F. Zhang, H. Wang, J. Wang, and K. Yang. Inexact primal–dual gradient projection methods for nonlinear optimization on convex set. *Optimization*, 69(10):2339–2365, 2020, https://doi.org/10.1080/02331934.2019.1696338.

ngHager2004 [70] H. Zhang and W. W. Hager. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optim.*, 14(4):1043–1056, 2004.

Dai2006 [71] B. Zhou, L. Gao, and Y.-H. Dai. Gradient methods with adaptive step-sizes. *Comput. Optim. Appl.*, 35(1):69–86, 2006.